You Won't Believe What They Just Said! The Effects of Political Deepfakes Embedded as Vox Populi on Social Media

Hameleers, M.; van der Meer, T.G.L.A.; Dobber, T.

DOI
10.1177/20563051221116346

Publication date
2022

Document Version
Final published version

Published in
Social Media + Society

Article

SM+S
social media + society

Social Media + Society
July-September 2022: 1–12

DOI: 10.1177/20563051221116346
journals.sagepub.com/home/sms
SAGE

You Won't Believe What They Just Said! The Effects of Political Deepfakes Embedded as Vox Populi on Social Media

Michael Hameleers, Toni G. L. A. van der Meer, and Tom Dobber

Abstract

Disinformation has been regarded as a key threat to democracy. Yet, we know little about the effects of different modalities of disinformation, or the impact of disinformation disseminated through (inauthentic) social media accounts of ordinary citizens. To test the effects of different forms of disinformation and their embedding, we conducted an experimental study in the Netherlands ($N = 1{,}244$). In this experiment, we investigated the effects of disinformation (contrasted to both similar and dissimilar authentic political speeches), the role of modality (textual manipulation versus a deepfake), and the disinformation's embedding on social media (absent, endorsed or discredited by an (in)authentic citizen). Our main findings indicate that deepfakes are less credible than authentic news on the same topic. Deepfakes are not more persuasive than textual disinformation. Although we did find that disinformation has effects on the perceived credibility and source evaluations of people who tend to agree with the stance of the disinformation's arguments, our findings suggest that the strong societal concerns on deepfakes' destabilizing impact on democracy are not completely justified.

Keywords

deepfakes, disinformation, endorsement, misinformation

Deepfakes—which we define as intentionally deceptive synthetic videos created with the use of Artificial Intelligence (AI) (e.g., Hancock & Bailenson, 2021; Westerlund, 2019)—have been regarded as a salient threat for democracy (e.g., Fletcher, 2018). Arguably, the affordances of deepfake techniques may result in the widespread application of realistic synthetic videos for cyber-bullying, propaganda, and blackmailing purposes (Maras & Alexandrou, 2019). Despite these concerns, however, we know very little about the effects of deepfakes compared to textual disinformation and authentic speeches (but see, for example, Lee & Shin, 2021). We know even less about how the deceptive embedding of deepfakes as "vox populi" on social media can amplify their impact. To arrive at a more comprehensive understanding of the consequences of deepfakes on social media, we created an original state-of-the-art political deepfake and systematically compared its effectiveness to authentic political speeches and textual disinformation.

We focus on disinformation in particular: The *intentional* manipulation, doctoring, or decontextualization of information to achieve a certain political goal (e.g., Bennett & Livingston, 2018; Freelon & Wells, 2020). Disinformation can involve different practices—ranging from the de-contextualization of information to the complete fabrication of false narratives (see, for example, Van der Linden et al., 2021). Yet, we know little about whether more cue-rich and realistic types of disinformation, such as deepfakes, are more effective in achieving the intended goals of disinformation than textual disinformation. Against this backdrop, we aim to explore the effects of political deepfakes on message credibility, issue agreement and source evaluations, and systematically compare their effectiveness to textual disinformation.

Disinformation's impact may be amplified by the affordances of social media. On social media, traditional journalistic gatekeepers can be circumvented, which empowers different malign actors to disseminate deceptive narratives while sidestepping verification and norms of balance and objectivity (e.g., Waisbord, 2018). In this article, we

University of Amsterdam, The Netherlands

Corresponding Author:
Michael Hameleers, University of Amsterdam, Nieuwe Achtergracht 166, Amsterdam, 1018 WV, Netherlands.
Email: m.hameleers@uva.nl

specifically focus on two aspects of disinformation that are relevant to consider in a social media context: The reliance on AI to make disinformation seem realistic (i.e., deepfakes) and the reliance on inauthentic source cues to mimic the many-to-many communication flows of social media (also see, for example, Zhang et al., 2013). Both of these factors may amplify the impact of disinformation by increasing its perceived realism and signaling social support.

The spread of disinformation has mainly been associated with radical right-wing issue positions, and the delegitimization of the established political order by foreign or domestic political actors that seek to amplify polarized divides (e.g., Bennett & Livingston, 2018; Marwick & Lewis, 2017). In this setting, this study will particularly look at disinformation containing explicit delegitimizing right-wing populist claims. Together, by integrating research on the role of modality and the social endorsement of right-wing populist disinformation online, this study aims to offer important new insights into the political consequences of participatory disinformation campaigns embedded as inauthentic coordinated behavior on social media.

Disinformation and Deepfakes: Textual versus Audiovisual Manipulation

Disinformation pertains to practices of manipulation, decontextualization, or fabrication to create a deliberately false, misleading, or harmful depiction of reality (e.g., Bennett & Livingston, 2018; Freelon & Wells, 2020; Marwick & Lewis, 2017). The dissemination and amplification of disinformation are for an important part afforded by social media (e.g., Bennett & Livingston, 2018; Freelon & Wells, 2020). Social media's user-to-user logic can amplify the reach of disinformation (Guess et al., 2019), and the absence of professional gatekeepers that ensure objectivity, civility, and balance also means that inaccurate and deceptive narratives can spread without being filtered out of people's newsfeeds (e.g., Bennett & Livingston, 2018). Social media may also empower the networked or participatory logic of disinformation campaigns (e.g., Starbird, 2019). By connecting political actors, mainstream media, ordinary citizens, and malign actors, social media can help to platform, mainstream, and amplify deceptive narratives—and reach vulnerable segments of the audience through the (micro)targeting of messages (e.g., Dobber et al., 2020).

We compare textual disinformation to deepfakes (also see Lee & Shin, 2021). Deepfakes can be regarded as synthetic videos created with the affordances of AI to make authentic (political) actors express inauthentic viewpoints (e.g., Paris & Donovan, 2020; Westerlund, 2019). Deepfakes can refer to image, voice, and video manipulations. In all cases, deep learning techniques are used to realistically manipulate and simulate the voice and/or (mouth) movements of real persons (Maras & Alexandrou, 2019). In this article, we focus on video-based deepfakes as these are arguably most prominent today (also see, for example, Dobber et al., 2020; Lee & Shin, 2021). Deepfakes can be created using Generative Adversarial Networks (GANs) (e.g., Yang et al., 2021)—which can be understood as an AI-driven collaboration between artificial neural networks that use large amounts of training materials (i.e., real footage of the depicted actor) to create realistic content resembling the original (also see Dan et al., 2021; Mirsky & Lee, 2021; Westerlund, 2019). In the context of our study, we focus on AI-generated deepfakes that are created using neural networks that have learned how to imitate the real person's facial expressions and gestures. Based on this learning process, the footage of an actual video of a politician is altered and integrated with a speech delivered by a voice actor.

Deepfakes are, by definition, intentionally deceptive (Dan et al., 2021; Hancock & Bailenson, 2021). Deepfakes are strategically crafted and targeted to harm and attack the depicted actor, for example, by making them express provocative, conflicting, or highly implausible claims (e.g., Dobber et al., 2020). Deepfakes may be harmful as developments in AI afford the generation of synthetic media that leave little trace of manipulation—which makes it difficult to identify deepfakes and instill resilience among recipients (e.g., Westerlund, 2019). Deepfakes, more than other modes of disinformation, offer a realism heuristic that should motivate heuristic processing and herewith circumvent the detection of deception by recipients (Sundar et al., 2021).

The Credibility of Disinformation versus Authentic Information

As people only have limited cognitive capacity to process all information systematically, they show a tendency to accept information as truthful rather than to doubt its veracity (Levine, 2014). Exploiting this truth-default or truth bias in information processing, disinformation may be perceived as credible and honest, especially when it stays close to reality (Hameleers et al., 2020). Yet, experimental research found that disinformation is still regarded as less credible than authentic information (e.g., Hameleers et al., 2020), which begs the question to what extent and how different forms of disinformation vis-à-vis authentic content are seen as credible.

In our experiment, we contrast disinformation with two speeches that were actually voiced by the depicted political actor. As a baseline comparison of the impact of deepfakes, we first of all contrast disinformation to an unrelated political speech on a different issue. Second, we contrast disinformation to a fragment of a speech that was actually expressed by the depicted political actor but deliberately cropped to emphasize a radical right-wing issue position resonating with the deceptive statements of the deepfake (a decontextualized message). Although this cropped speech may not be regarded as disinformation in the strictest sense as the material is not manipulated or fabricated, the creation of the isolated and decontextualized fragment could be regarded as

intentionally deceptive mal-information (Wardle & Derakhshan, 2017). We will refer to this decontextualized use of real information as malinformation in the remainder of this article. In the context of our study, malinformation refers to real information that is deliberately decontextualized to make the political actor seem more aligned with radical right-wing issue positions than he actually is.

We aimed to use realistic manipulations resonating with the right-wing conservative views of the depicted politician, but at the same time deviated from factually accurate information and familiar political statements to reflect a delegitimizing disinformation campaign. Hence, we are interested in what happens if, via the manipulation of text or video, political actors are made to express more extreme viewpoints than they would typically voice. As deepfakes are made to attack someone or a certain group (a political actor in this case), it is likely that the statements used are significantly more extreme and less familiar for recipients. We therefore believe that the deviation from reality is externally valid. Considering that the realism index offered in videos can be convincing, deepfakes may be able to let people believe anything the manipulated video shows just because it looks authentic.

At the same time, to maintain a connection to reality, we did not manipulate political positions that would run counter to the politician's political orientation. Distinguishing between malinformation, disinformation, and unrelated authentic speeches, we first of all ask to what extent participants can separate disinformation from the two authentic speeches in terms of their credibility ratings: What is the difference in the perceived credibility of authentic information, malinformation and (synthetically) manipulated disinformation ? (RQ_1).

Disinformation's Effects on Credibility, Issue Agreement and Support for the Depicted Politician

In this article, we compare the effects of deepfakes to textual disinformation. The specific qualities of audiovisual information used in synthetic deepfake videos may add persuasive power to disinformation campaigns as audiovisual materials offer a stronger and seemingly less filtered link to reality than texts (Messaris & Abraham, 2001; Powell et al., 2018). Therefore, news consumers may be less likely to doubt the veracity of audiovisual information than textual information. Hence, deepfakes have been regarded as dangerous as the manner of manipulation results in seemingly authentic depictions of real persons (Maras & Alexandrou, 2019). Disinformation videos may herewith offer a realism heuristic (Sundar et al., 2021). As a consequence, people may be less likely to systematically process the arguments of disinformation because the audiovisual mode of presentation signifies truthfulness. Against this backdrop, deepfakes may not prime suspicion as recipients are not likely to doubt the authenticity of AI-generated media, and rather process the message heuristically driven by the index of realism that is offered.

Extant literature found some support for the premise that deepfakes are more effective than other modes of disinformation. Sundar et al. (2021) found that video-based disinformation was seen as more realistic than text—or audio-based disinformation, but only among lower involved participants. The conclusions of Lee and Shin (2021) and Hwang et al. (2021) point in the same direction: Deepfake videos are seen as slightly more vivid and credible than textual disinformation, but the differences are relatively small. While Lee and Shin (2021) focused on the controversial issues of abortion and marijuana legalization and Hwang et al. (2021) on a deceptive claim featuring the CEO of Facebook, we focus on the effects of disinformation delegitimizing an established political actor. We make this actor voice the radical right-wing issue position that immigrants pose a threat to the native people—a deceptive statement that is frequently present in disinformation campaigns disseminated by right-wing populists (e.g., Bennett & Livingston, 2018). Such messages may be disseminated by foreign actors or domestic political opponents to delegitimize established politicians in Western democracies, increase polarized divides, and raise cynicism (e.g., Bennett & Livingston, 2018; Lukito et al., 2020). These right-wing populist messages emphasizing sociopolitical cleavages, conflict, distrust, and identity politics resonate with disinformation's aim to increase cleavages and sow discord (Marwick & Lewis, 2017). Against this backdrop, we focus on the polarizing and politicized issue of anti-immigration as it strongly reflects the disinformation logic in Western democracies.

We specifically look at three dependent variables to assess the impact of disinformation: Its perceived credibility, issue-agreement with disinformed statements, and the rating of the depicted politician. We selected these dependent variables for different reasons. From the perspective of increased concerns about disinformation's impact, it is important to assess whether manipulated information is seen as equally credible as authentic information (e.g., Lee & Shin, 2021) and whether intentional attacks on prominent political figures can harm people's political evaluations and agreement with the issue at hand (e.g., Dobber et al., 2020). Together, these three variables can be used to assess the effectiveness of the delegitimizing goals pursued by disinformation actors: To make receivers accept inauthentic content, steer the political evaluations of news users in line with the manipulated message, and/or harm the credibility and liking of the depicted politician (e.g., Bennett & Livingston, 2018; Dobber et al., 2020).

We postulate the following hypotheses on the relative effectiveness of deepfakes: Exposure to a deepfake has a stronger effect on the perceived credibility of disinformation (H1a), issue agreement with statements made in disinformation (H1b), and negative evaluations of the depicted politician (H1c) than exposure to textual disinformation.

Motivated Reasoning and Disinformation's Persuasiveness

In line with the premises of defensive motivated reasoning, people may process information in a way to confirm and reassure their prior beliefs while avoiding or rejecting discrepant views that challenge their identities and beliefs (e.g., Kunda, 1990). Based on this, we expect that people with stronger preexisting anti-immigration beliefs and more extreme right-wing political orientations are most likely to perceive disinformation as credible because such statements reassure their prior beliefs (also see Schaewitz et al., 2020). We additionally expect that their preexisting negative schemata and stereotypes toward immigrants—in this paper measured as anti-immigration beliefs—are primed or activated when being exposed to the manipulated political speech that cultivates a clear anti-immigration stance.

When studying motivated reasoning, we focus on the extremity of ideological orientations and anti-immigration beliefs as these evaluations tap into the resonance of the manipulated messages on both an ideological and attitudinal level: The more people identify with the right-wing fringe of the ideological spectrum, and the more they oppose immigration, the stronger the congruence between radical right-wing disinformation and their existing worldviews should be. Effects based on attitude-congruent exposure can be understood as driven by a confirmation bias. In line with this, we expect that disinformation has the strongest effects when it confirms people's prior beliefs and ideological orientations (also see, for example, Knobloch-Westerwick et al., 2017). We therefore hypothesize: Political disinformation containing right-wing populist interpretations has the strongest effects on credibility and issue agreement among participants with more extreme-right-wing orientations (H2a) and stronger anti-immigration beliefs (H2b).

We finally look at the delegitimizing impact of exposure to disinformation targeted at supporters of the depicted politician. Dobber et al. (2020) found some support for such a harmful impact of deepfakes: When synthetic videos were microtargeted to participants supporting the depicted politician and his values, deepfakes harmed the positive evaluation of the politician. As an important aim of disinformation campaigns is to sow discord and delegitimize established voices (e.g., Bennett & Livingston, 2018; Marwick & Lewis, 2017), targeted deepfakes may succeed in harming support for established politicians by reaching their constituents with debasing content. We therefore formulate the following hypothesis: Exposure to political disinformation mostly harms the rating of the depicted politician among participants supporting the politician and his viewpoints (H2c).

The Role of Disinformation's Embedding: Social Endorsement and Discrediting Cues

The affordances of social media can be used to make inauthentic content seem like the online exchanges between opiniated ordinary citizens. In this regard, digital disinformation is often paired with a social embedding, which is either authentic (a social media user that communicates an issue position) or inauthentic (a bot or troll that mimics an authentic citizen) (Shao et al., 2018; Starbird, 2019; Zhang et al., 2013).

In this paper, we distinguish between two alternative ways in which disinformation can be embedded as "vox populi" (Lukito et al., 2020) on social media: (1) to discredit the depicted political actor by making him or her look incredible, dishonest, or corrupt and (2) to endorse the political actor by claiming political success—hereby amplifying support for this actor. We expect that these embeddings amplify the effects of disinformation as they add more familiarity and authenticity to such content—which is in line with the so-called "illusory truth effect" (e.g., Hasher et al., 1977). Extending this argument, we postulate that embedding disinformation in social media formats also creates an illusion of truthfulness by making disinformation seem like the social media interactions people are exposed to on a daily basis.

Extant research on the effects of related right-wing populist communication found that using references to ordinary citizen sources can enhance the effectiveness of such arguments (Hameleers & Schmuck, 2017). This can be explained as an in-group serving bias (Taylor & Doria, 1981): People are more likely to uncritically accept a political message when it comes from members of the social group they belong to. In this study, we specifically expect that when disinformation is allegedly communicated by an ordinary citizen that either discredits or endorses the message, it may be seen as more credible than when such a reference is absent. The ordinary citizen cue thus creates an illusion of truth or a realism heuristic for disinformation, and signifies similarity with recipients.

For the other two dependent variables—issue agreement and source evaluations—the direction of the embedding should matter: Issue agreement should be stronger when the social media cue endorses the message and weaker when the political statements are discredited. For the evaluations of the depicted politician, discrediting the politician should result in more negative source evaluations, whereas praising his speech should cultivate more support. We therefore raise the following hypotheses: Disinformation embedded on social media has a stronger effect on message credibility than disinformation without an embedding (H3a). Social media cues that endorse the message result in higher levels of issue agreement (H3b) and more positive source evaluations (H3c) than the absence of an endorsement. Discrediting cues result

Table 1. Overview of Experimental Conditions.

(Dis)information Embedding	Authentic Unrelated	Malinformation	Disinformation
No embedding	(1) Unrelated authentic speech, no embedding (an authentic fragment of a political speech on the country' s progress)	(2) Decontextualized authentic speech, no embedding (the message leaves out the context of right-wing populist statements on immigration that were expressed by the political actor)	(3) Disinformation, no embedding (the depicted political actor is made to voice radical right-wing hate speech targeted at immigrants)
Endorsed		(4) The authentic message is embedded in the social media feed of an ordinary citizen who approves of the right-wing populist statements of the political actor (i.e., saying it is a strong statement)	(5) The disinfolmation message is embedded in the social media feed of an ordinary citizen who approves of the hate speech (i.e., saying it is a strong statement)
Discredited		(6) The authentic message is embedded in the social media feed of an ordinary citizen who disapproves of the right-wing populist statements of the political actor (i.e., saying it is a weak statement)	(7) The disinfolmation message is embedded in the social media feed of an ordinary citizen who disapproves of the hate speech (i.e., saying it is a weak statement)

Note. All seven conditions were either presented as textual or as audiovisual information (a deepfake for the disinformation conditions).

in lower issue agreement (H3d) and more negative source evaluations (H3e) than the absence of a discrediting cue.

Method

Design and Stimuli

The experiment is set up with the following between-subjects factorial design: 3 (Disinformation: unrelated authentic information versus malinformation versus disinformation) × 3 (Embedding: absent versus endorsed versus discredited × 2 (Modalities: textual versus audio-visual). The true factorial design corresponds to 18 conditions, but we excluded the embedding factor for the unrelated authentic political speech. There were thus 14 different conditions that participants were randomly assigned to. Table 1 depicts an overview of conditions and the difference between authentic unrelated and malinformation messages in particular.

To create the stimuli, we used a real speech of the depicted Dutch politician (Sybrand van Haersma Buma). This politician is the former leader of the Dutch Christian-Democrats and known for rather conservative views on issues such as immigration and national politics (see more information on the profile and viewpoints of this political actor in Supplemental Appendix A). In the manipulations, we had to deviate from true statements as we wanted to mimic the delegitimizing narratives of disinformation and thus made someone express more extreme viewpoints than normally expressed by this person (also see Dobber et al., 2020). The manipulations were the result of attempting to strike a balance between plausibility (i.e., not changing the political orientation of the politician) and extremity (i.e., making the actor look bad by adding an extremist component to statements). Crucially, the politician used in the disinformation conditions is not considered a right-wing populist, but his overall ideology and rhetoric can be manipulated and placed out of context to make his existing conservative viewpoints more extreme.

The low salient lecture used as a source of the manipulations was given in 2017, and the recording is available online, but Dutch citizens rarely view these recordings (we controlled for prior exposure in the experiment). From this long video, we selected two short fragments (about 45 s): One that was unrelated to the political deepfake we intended to create (about technological and societal progress and upward mobility), and one that was more closely related to the right-wing populist statements we fabricated (about immigrants who stick to the language, religion, and culture of their country of origin—similar to the content of the deepfake's arguments). We refer to this condition as malinformation. We transcribed the videos in verbatim, which formed the basis of the textual reference materials that were presented as news items (see Appendix D for the political speeches).

For the deepfake, we hired an external high-skilled VFX and AI artist known to have made extremely credible deepfakes in the past. The artist used the control conditions as reference and training materials, which also enhanced the similarity of the deepfake to the authentic videos used in the experiment. We additionally hired a professional voice actor to imitate the depicted politician. This voice recording was used by the visual artist to create the deepfake (i.e., mouth manipulation). After different rounds of editing and weeks of

training, a final version of the deepfake was completed (stills are included in Appendix C). In the disinformation conditions, the politician ostensibly stated that immigrants are enabled to increasingly influence our country's traditions and that people from backward societies are more often inclined to commit violent crimes such as rapes and robberies (see Appendix D).

In the testing and development phase, the deepfake was rated as very credible and similar to authentic speeches. Pilot testing revealed that people who did not know the intentions of the experiment found it hard to distinguish the deepfake from an authentic video (there were no significant mean differences in credibility ratings). Specifically, the deepfake was rated as almost equally credible (M=3.81, SD=1.54) as malinformation (M=3.99, SD=1.34). Due to ethical concerns and the risk of distributing a deepfake outside the experimental (controlled) setting, the deepfake video is not included in this article.

We used subtitles in the authentic and deepfake videos to make sure that the meaning was decoded adequately by participants. We closely matched the authentic (de-contextualized) videos and deepfakes: They were equal in length (45 s), similar in quality (HD), had exactly the same neutral background and source cues, and used the same camera angle to depict the politician.

For the textual conditions, respondents were shown a news-like article about a statement recently made by the politician. These manipulations consisted of short articles that introduced a quote coming from the politician. Three versions of the text were created that resembled the statements made in audiovisual material of the (1) unrelated political speech, (2) malinformation, and (3) the deepfake. See Appendix D for the full texts of the stimuli.

To manipulate the social media embedding of the (dis)information, we presented the videos and texts as if they were shared online, and commented on by an ordinary citizen (or a troll in case this reflects coordinated inauthentic behavior). A Twitter profile was faked with a randomly (AI) generated profile photo of a middle-aged White male. First, in the endorsement conditions, this fake profile endorsed the statements by stating that it was a strong statement and that the politician was "someone who finally had the guts to say what everyone is thinking." In the discrediting embedding condition, the statement was discredited by the fake person who argued that it is "unbelievable that a mainstream politician would spread fact-free radical-right wing hate speech" (also see Table 1). Appendix F includes an example of the endorsement and discrediting condition. Appendix I reports on the outcomes of pretreatment robustness checks and manipulation checks, which ensured that the manipulations were perceived as intended.

Sample

Data collection was outsourced to a large international research agency (Kantar Lightspeed) who relied on large databases of panelists from mixed resources (originally recruited via telephone, e-mail, and offline recruiting). We enforced quotas to ensure a distribution of age, gender, and education that matched the composition of the Dutch population (recent census data were used for matching). The mean age of participants was 49.76 years (SD=14.66); 21.3% was lower educated and 31.2% completed a higher level of formal education (47.5% moderate). In our final sample, females were slightly overrepresented with 55.6%. Looking at indicators of ideological preferences, our sample was relatively balanced (41.3% left-wing, 47.4% right-wing, 11.3% don't know/prefer not to say). This balance is relevant as we use extreme right-wing orientations as a moderator to assess whether an ideological resonance between the (dis)information and the preexisting ideological beliefs makes disinformation more persuasive. Finally, 46.1% at least was somewhat interested in politics (the low and high extremes of the scale were occupied by 6.8% and 6.8%, respectively). Our final sample size was 1,244—and we achieved a completion rate of 85.6% (the proportion of invited participants that completed the full survey). The approximate sample size was predetermined based on power analyses (the tool G*Power was used, see Faul et al., 2009). Based on previous experiments measuring the effects of disinformation (e.g., Schaewitz et al., 2020), we expected relatively small (indirect) effect sizes (<.20). Our dependent variables were also used in previous experiments, which helped us to determine an appropriate sample size to obtain a power of .80 (80+ participants per group).

Dependent Variables

We measured the effects of disinformation by looking at three different dependent variables: Agreement with the ideational stance and arguments of the (dis)information (issue agreement), the (in)credibility/perceived authenticity of the message (credibility), and negative evaluations of the political actor depicted in the message (evaluations of the depicted politician). Details on item formulation and scale construction can be found in Appendix G.

Moderators

We established the resonance between the disinformation's statements and prior ideological orientation and attitudes by measuring anti-immigration beliefs and more extreme right-wing political orientations. Both scales were measured before exposing participants to the treatment. To measure extreme right-wing orientations, we used the classical left-right ideological orientation scale that we recoded into a binary variable discriminating between more extreme right-wing orientations versus other ideological orientations (multiple robustness checks with alternative re-coding procedures were considered). We measured prior levels of support/approval of the depicted political actor Buma using a slider ranging from 0 (*no support at all*) to 100 (*full support*). Details on item formulation and scale construction can be found in Appendix G.

Procedures

Participants entered the online experiment on a non-mobile device via an e-mail invitation sent by the research agency. First, they were informed about the study and the type of questions they would respond to. The experiment received ethical approval from the University's ethical committee (University of Amsterdam, reference number 2022-PCJ-14411) that ensured that all informed consent, debriefing, and anonymity concerns were dealt with extensively (in line with AoIR guidelines). Second, participants completed demographic questions, followed by measures for our moderators. In the next block, they were forwarded to the experimental module. After receiving detailed instructions about the (audiovisual nature of) the political speech (i.e., turn your audio on, pay attention to the message), they saw a political speech that matched their experimental condition (randomized). All speeches were equal in length (less than 1 min) and we gave a similar minimum exposure time for the textual conditions to enhance comparability. After the treatment, participants were forwarded to the post-treatment survey. Here, they answered questions about the credibility and authenticity of the (dis)information, their agreement with political statements that matched the content of the (dis) information, and their evaluations of the depicted politician. More details on the ethical considerations and procedures can be found in Appendix B.

Results

The Credibility of (Deep)fakes versus Authentic Information

We first of all assessed the differences in perceived credibility between the (a) authentic unrelated messages, (b) the malinformation messages, and the (c) disinformation messages (RQ_1). We rely on a one-way analysis of variance with Bonferroni-corrected pairwise mean score comparisons to map these differences (video-based and textual stimuli are combined). For these analyses, we use the conditions variable (disinformation versus malinformation versus authentic unrelated information) as the independent variable and the credibility rating scale as the dependent variable. We see a significant difference across the messages that deviate from facticity to different extents, $F(2, 1{,}241) = 34.10$, $p < .001$, partial $\eta^2 = .052$. More specifically, unrelated authentic information is rated as significantly more credible ($M = 4.35$, $SD = 1.08$) than malinformation ($M = 3.81$, $SD = 1.11$) or disinformation ($M = 3.47$, $SD = 1.37$). Although the mean difference between disinformation and malinformation is significant ($\Delta M = .34$, $SE = .07$, $p < .001$), it is less strong than the difference between the unrelated authentic and malinformation message ($\Delta M = .54$, $SE = .10$, $p < .001$) or the difference between the unrelated authentic message and disinformation ($\Delta M = .88$, $SE = .11$, $p < .001$). Taken together, disinformation is seen as less credible than authentic information, and malinformation is rated as substantially less credible than an authentic speech that is unrelated to the disinformation's claims (also see Figure 1 for a comparison of mean credibility scores across conditions).

The Added Persuasive Power of Deepfakes versus Textual Disinformation

We expected that deepfakes have a stronger effect on the perceived credibility (H1a), issue agreement (H1b), and negative evaluations of the depicted politician (H1c) than textual disinformation. Based on the (corrected) mean score comparisons, we can conclude that a deepfake is not rated as significantly more credible ($M = 3.41$, $SD = 1.42$) than an equivalent textual disinformation message, $M = 3.54$, $SD = 1.30$; $t(539) = 1.06$, $p = .291$. Although deepfakes generate slightly more agreement ($M = 4.56$, $SD = 1.61$) than textual disinformation ($M = 4.49$, $SD = 1.62$), the difference is again not significant, $t(539) = -.50$, $p = .615$. Regarding participants' evaluations of the depicted politician, we also do not find support for a difference between exposure to textual disinformation ($M = 3.97$, $SD = 1.07$) and the deepfake ($M = 3.96$, $SD = 1.12$). In other words, we do not find added persuasive power of a deepfake compared to textual disinformation.

Using ordinary least square (OLS) regression models (see Appendix H for full regression tables and notes on the analysis strategies that were used), we compare the effects of authentic versus disinformation across the two modalities. Before testing the hypotheses, we have conducted a series of robustness checks that are reported in Appendix H. Turning to our hypotheses, we first of all see a negative, significant two-way interaction effect between disinformation exposure (versus authentic information) and modality (text versus video) on perceived credibility ($B = -.47$, $SE = .14$, $\beta = -.16$, $p = .001$). Contrary to H1a, this means that a deepfake was perceived *less* credible than textual disinformation. We do not find a significant two-way interaction effect between modality and disinformation on issue agreement ($B = -.05$, $SE = .18$, $\beta = -.01$, $p = .809$). This does not support H1b. We also fail to find support for H1c, as the two-way interaction effect between disinformation exposure and modality on evaluations of the depicted politician is not significant ($B = .16$, $SE = .12$, $\beta = .07$, $p = .159$). Taken together, our findings do not offer support for the expectation that exposure to a political deepfake has stronger effects than textual disinformation.

The Moderating Effect of Extreme Right-wing Orientations and Anti-immigration Beliefs

We finally look at the interaction effects between disinformation exposure and prior attitudinal biases aligned with the arguments voiced in the fabricated disinformation

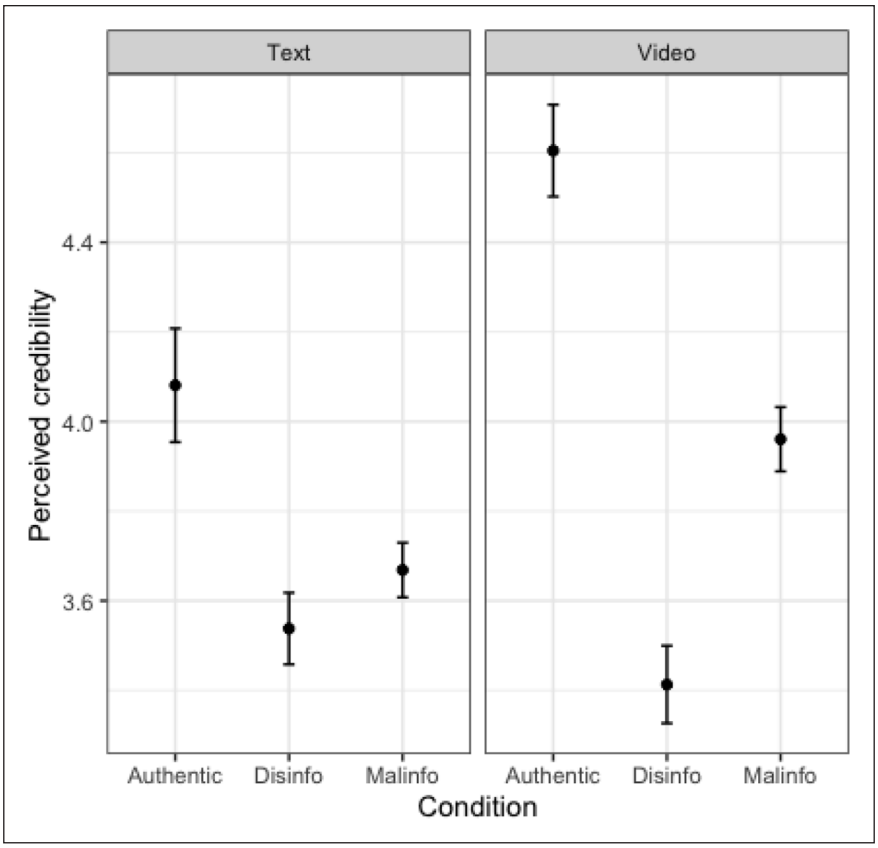

Figure 1. The perceived credibility of disinformation, malinformation, and unrelated authentic information across textual and audiovisual modalities.

message: More extreme right-wing political orientations (H2a) and anti-immigration beliefs (H2b). Plots are presented in Figure 2. For perceived credibility, we see no support for a significant two-way interaction effect between disinformation exposure and more extreme right-wing political orientations ($B = .21$, $SE = .18$, $\beta = .05$, $p = .246$)—which does not offer support for H2a. However, we do find support for H2b: The stronger people's anti-immigration beliefs, the more likely they perceive disinformation as credible ($B = .23$, $SE = .05$, $\beta = .43$, $p = < .001$). We do not find a similar effect for the interaction effect between attitudinal congruence and disinformation exposure on post-treatment issue agreement ($B = .00$, $SE = .03$, $\beta = -.00$, $p = .991$).

The negative two-way interaction effect between anti-immigration beliefs and disinformation exposure on negative evaluations of the politician ($B = -.11$, $SE = .04$, $\beta = -.25$, $p = < .001$) indicates that when disinformation resonated with prior beliefs, people were more likely to positively evaluate the depicted politician making the claims they supported. A similar effect was found for more extreme right-wing political orientations ($B = -.30$, $SE = .15$, $\beta = -.08$, $p = .046$). People thus rated the political actor more positively when disinformation was in line with their existing views.

To more specifically test the assumptions underlying H2c, we ran a regression model estimating the interaction effect between existing support for the depicted politician and exposure to (audiovisual) disinformation. The results are nonsignificant, both for the two-way interaction effect between prior support and exposure to disinformation on the rating of the depicted politician ($B = .00$, $SE = .00$, $\beta = .01$, $p = .870$) and the three-way interaction effect in which we specify these effects for a deepfake versus textual disinformation ($B = .00$, $SE = .00$, $\beta = .05$, $p = .298$). In line with these findings, it can be concluded that exposure to disinformation—either in textual form or presented as a deepfake—does not harm the evaluations of the depicted politician among people more inclined to support or approve of this political actor in the first place. There is thus no support for H2c. Although participants with stronger anti-immigration beliefs and more extreme right-wing orientations may become more supportive of a politician when he voices statements that resonate with their views, we find no support for a delegitimizing impact among supporters of the politician (H2c).

Finally, in Model V of the OLS regressions, we assessed the three-way interaction effects between exposure to disinformation, modality, and issue congruence (more extreme right-wing political orientations and anti-immigration beliefs)

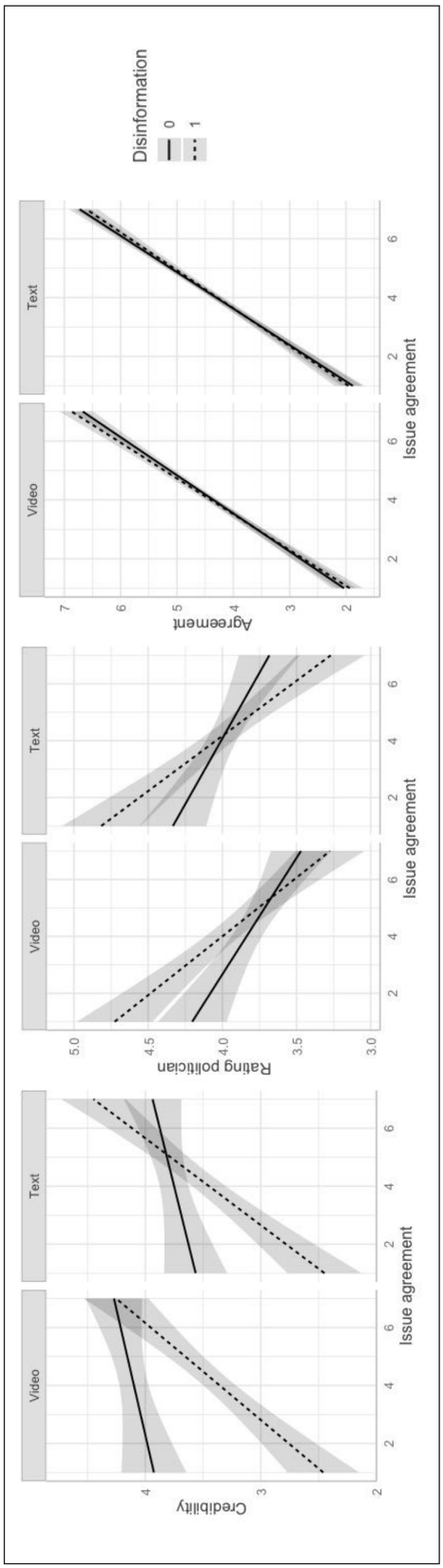

Figure 2. The effects of textual disinformation and deepfakes on credibility (left), rating of depicted politician (mid), and prior issue agreement (right) moderated by prior levels of agreement with disinformation's statements (anti-immigration beliefs).

to test whether deepfakes (compared to textual modes of [dis] information) are most persuasive for participants with congruent prior attitudes. We only find a significant three-way interaction effect for deepfake exposure and more extreme right-wing political orientations on perceived credibility, but the effect is negative ($B=-.62$, $SE=.27$, $\beta=-.10$, $p=.023$). Contrary to what we expected, deepfakes are not more effective in triggering credibility, issue agreement, or evaluations of the depicted politician among participants aligned most with the attitudinal stance of the manipulated information.

The Impact of Socially Embedded (Deep)Fakes

We expected that, irrespective of the direction of the embedding, credibility increases when a (fake) ordinary citizen is used to embed the message (H3a). In line with this, we find a significant two-way interaction effect between exposure to disinformation and a discrediting cue ($B=.32$, $SE=.16$, $\beta=.09$, $p=.047$) on message credibility. This supports H3a. However, we do not find support for a significant two-way interaction effect between disinformation exposure and an endorsing cue ($B=.21$, $SE=.17$, $\beta=.06$, $p=.205$). H3a is thus partially supported: When disinformation is accompanied by a social media cue that stresses that the statements are weak and nonsensical, participants are more likely to regard this message as credible. However, a social media cue that offers support for the message by emphasizing that the disinformed statements are strong does not increase the credibility of the message.

We do not find significant two-way interaction effects for exposure to endorsing ($B=-.13$, $SE=.22$, $\beta=-.03$, $p=.544$) or discrediting ($B=-.33$, $SE=.22$, $\beta=-.07$, $p=.134$) cues and disinformation on issue agreement—which offers no support for H3b. The two-way interaction effect between endorsement and disinformation exposure on the rating of the depicted politician is significant ($B=-.28$, $SE=.14$, $\beta=-.10$, $p=.048$), whereas we do not find this for the interaction effect between discrediting cues and disinformation exposure ($B=-.15$, $SE=.14$, $\beta=-.05$, $p=.285$). Endorsing a disinformation message results in a less negative evaluation of the depicted politician, which is in line with H3c. Finally, the three-way interaction effects between exposure to disinformation, the video cue, and an endorsement or discrediting were not significant: The embedding strategy of disinformation campaigns on social media has only a very limited impact on credibility, and its impact is not stronger for textual disinformation than for deepfakes.

Discussion

Although previous research has experimentally tested the impact of political deepfakes (Dobber et al., 2020; Vaccari & Chadwick, 2020), important questions remain: Do deepfakes in the political realm have a persuasive advantage over textual modes of deception, and are its effects contingent upon the endorsement by (fake) social media accounts in a

"participatory" disinformation order (Lukito et al., 2020; Starbird, 2019)? Our main findings indicate that disinformation was rated as substantially less credible than an unrelated authentic speech. However, disinformation was not rated as substantially less credible than malinformation based on an authentic speech of the depicted political actor. Finally, exposure to a deepfake did not yield stronger effects than exposure to textual disinformation.

These findings contradict the ubiquitous concerns on deepfakes in the current digital information age and are not in line with literature on the persuasiveness of multimodal framing (Powell et al., 2018) or deepfakes (Lee & Shin, 2021). To some extent, the lack of effects is in line with previous research on deepfakes indicating that deepfakes do not directly mislead news users (Dobber et al., 2020; Vaccari & Chadwick, 2020). Rather, deepfakes may have a more indirect effect by making recipients unsure on what or whom to believe, which, in turn, reduces people's trust in (online) news (Vaccari & Chadwick, 2020).

Regarding the embedding of disinformation on social media, we found that discrediting the fabricated statements can make disinformation appear more credible. The social endorsement cue does not have this effect. However, endorsing a disinformation message on social media resulted in a more positive evaluation of the depicted politician compared to the absence of such an endorsement. This can be understood as an in-group serving bias: The presence of a source cue similar to the recipient may enhance credibility. These findings show that disinformation agents employing trolls or bots that use inauthentic social media profiles are only effective in increasing the message's credibility when a fake message is *dis*credited. The democratic implications of these findings are optimistic: Deepfakes, at least based on the current state-of-the-art, do not seem to be as dangerous for society as assumed (Paris & Donovan, 2020). While concerns about information pollution and eroding public trust remain, deepfakes' ability to destabilize democracy should not be overstated.

We did find some support for the conditional effects of deepfakes and textual disinformation. Overall, disinformation is more effective in affecting the credibility ratings and positive evaluations of the depicted politician among news users already inclined to support the attitudinal stance of disinformation's statements—which is in line with extant research on the indirect impact of disinformation campaigns (e.g., Schaewitz et al., 2020). We can understand this as a confirmation bias (e.g., Knobloch-Westerwick et al., 2017): Disinformation that is congruent with people's prior beliefs may reinforce these existing beliefs. Yet, we did not find support for a moderating role of prior levels of issue agreement on disinformation's impact on message congruent beliefs. We can explain this as a ceiling effect: People already inclined to support the attitudinal stance of disinformation are not further bolstering their beliefs based on one single message that reassures the attitudes they already hold.

Despite contributing to our understanding of the political consequences of exposure to socially endorsed deepfakes, this article has a number of limitations. As the deepfake was almost as credible as an authentic decontextualized video (malinformation), we believe that the lack of effects was not only due to technological failures of the deepfake itself but also the lack of plausibility of the extremist political statements associated with the more moderate political actor. Yet, we aimed to strike a balance between actually voiced statements by the political actor and a delegitimizing narrative that would harm the political actor by making him look bad. To achieve this, some deviation from the truth and familiar statements was needed. The finding that people do not clearly differentiate between fabricated disinformation and decontextualized malinformation is an important finding in its own right: In times when the truth has become more relative (e.g., Van Aelst et al., 2017), people may also distrust authentic information when it triggers suspicion due to its unusual nature.

This specific trade-off between audiovisual credibility and argumentative discrepancies needs to be teased out further in the future: How far can a deepfake deviate from a political actor's profile to still be perceived as credible, and what persuasive techniques can be used to make inauthentic arguments seem real? Hence, future research may experiment with different conditions that are more or less plausible and more or less distant to the everyday communication of a known target. They may also more centrally take into account people's existing knowledge and beliefs related to the depicted politician's issue positions. If deepfakes are no longer credible when they deviate too much from reality, this may have positive implications for democracy: There are limits to the "fake reality" shown in synthetic videos, and deepfakes are not capable of making everyone say anything while remaining credible.

We should also note that the construct of perceived credibility we used may mean different things for different participants. While some may interpret the statements as referring to the authenticity of the presented materials, others may see it as the "truth value" of the statements themselves (Lewandowsky, 2021). Against this backdrop, some participants may find a deepfake uncredible because the statements do not have truth value, whereas others detect deception in the presentation of the video. Although robustness checks distinguishing between these drivers of credibility do not point to substantial differences, we suggest future research to rely on a more comprehensive multidimensional measure of credibility that distinguishes between these interpretations. In addition, although exposure to one short deepfake on its own may not affect polarization or political evaluations, the cumulative (targeted or algorithmic) exposure to attitude-consistent disinformation may, over time, exert a stronger influence on people's beliefs and behaviors. Finally, future research may also need to take individual-level differences into account that could predict susceptibility and resilience to disinformation, such as people's trust in social versus mainstream media, and formats more likely to contain disinformation.

As a key take-away point, we stress that although the disrupting impact of deepfakes on democracy should not be overstated, deepfakes' ability to become part of native online political discussions may offer a persuasive advantage when it can find nuanced ways to delegitimize political actors or amplify the political beliefs of targeted groups in society via social media.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: An unrestricted gift stemming from the 2020 Content Policy Research Initiative at Facebook.

ORCID iD

Michael Hameleers https://orcid.org/0000-0002-8038-5005

Supplemental Material

Supplemental material for this article is available online.

References

Bennett, L. W., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), 122–139. https://doi.org/10.1177/0267323118760317

Dan, V., Paris, B., Donovan, J., Hameleers, M., Roozenbeek, J., van der Linden, S., & von Sikorski, C. (2021). Visual mis- and disinformation, social media, and democracy. *Journalism & Mass Communication Quarterly*, *98*(3), 641–664. https://doi.org/10.1177/10776990211035395

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. H. (2020). Do (microtargeted) deepfakes have real effects on political attitudes? *International Journal of Press/Politics*, *26*(1), 69–91. https://doi.org/10.1177/1940161220944364

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Fletcher, J. (2018). Deepfakes, Artificial Intelligence, and some kind of dystopia: The new faces of online post-fact performance. *Theatre Journal*, *70*(4), 455–471. https://doi.org/10.1353/tj.2018.0097

Freelon, D., & Wells, C. (2020). Disinformation as political communication. *Political Communication*, *37*, 145–156. https://doi.org/10.1080/10584609.2020.1723755

Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, *5*(1), 1–8. https://doi.org/10.1126/sciadv.aau4586

Hameleers, M., Powell, T. E., van der Meer, G. L. A., & Bos, L. (2020). A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. *Political Communication*, *37*, 281–301. https://doi.org/10.1080/10584609.2019.1674979

Hameleers, M., & Schmuck, D. (2017). It's us against them: A comparative experiment on the effects of populist messages communicated via social media. *Information, Communication & Society*, *20*(9), 1425–1444. https://doi.org/10.1080/1369118X.2017.1328523

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, Behavior and Social Networking*, *23*(4), 149–152. http://doi.org/10.1089/cyber.2021.29208.jth

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, *16*, 107–112.

Hwang, Y., Ryu, J. Y., & Jeong, S. H. (2021). Effects of disinformation using deepfake: The protective effect of media literacy education. *Cyberpsychology, Behavior, and Social Networking*, *24*(3), 188–193. https://doi.org/10.1089/cyber.2020.0174

Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2017). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Communication Research*, *47*, 104–124. https://doi.org/10.1177/0093650217719596

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Lee, J., & Shin, S.-Y. (2021). Something that they never said: Multimodal disinformation and source vividness in understanding the power of AI-enabled deepfake news. *Media Psychology*, *25*, 531–546. https://doi.org/10.1080/15213269.2021.2007489

Levine, T. R. (2014). Truth-Default Theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, *33*(4), 378–392. https://doi.org/10.1177/0261927X14535916

Lewandowsky, S. (2021). Conspiracist cognition: Chaos, convenience, and cause for concern. *Journal for Cultural Research*, *25*(1), 12–35. https://doi.org/10.1080/14797585.2021.1886423

Lukito, J., Suk, J., Zhang, Y., Doroshenko, L., Kim, S. J., Su, M.-H., Xia, Y., Freelon, D., & Wells, C. (2020). The wolves in sheep's clothing: How Russia's Internet Research Agency tweets appeared in U.S. news as vox populi. *The International Journal of Press/politics*, *25*(2), 196–216. https://doi.org/10.1177/1940161219895215

Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *International Journal of Evidence & Proof*, *23*(3), 255–262. https://doi.org/10.1177/1365712718807226

Marwick, A., & Lewis, R. (2017, May 15). *Media manipulation and disinformation online*. Data & Society Research Institute. https://datasociety.net/output/media-manipulation-and-disinfo-online/

Messaris, P., & Abraham, L. (2001). The role of images in framing news stories. In S. D. Reese, O. H. Gandy, & A. E. Grant (Eds.), *Framing public life* (pp. 215–226). Erlbaum.

Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys (CSUR)*, *54*(1), 1–41.

Paris, B., & Donovan, J. (2020). Deepfakes and cheapfakes: The manipulation of audio and visual evidence. *Data & Society Report*. https://datasociety.net/library/deepfakes-and-cheapfakes/

Powell, T. E., Boomgaarden, H. G., De Swert, K., & de Vreese, C. H. (2018). Video killed the news article? Comparing multimodal framing effects in news videos and articles. *Journal of*

Broadcasting & Electronic Media, *62*(4), 578–596. https://doi.org/10.1080/08838151.2018.1483935

Schaewitz, L., Kluck, J. P., Klösters, L., & Krämer, N. C. (2020). When is disinformation (in) credible? Experimental findings on message characteristics and individual differences. *Mass Communication & Society*, *23*, 484–509. https://doi.org/10.1080/15205436.2020.1716983

Shao, C., Ciampaglia, G., Varol, O., Yang, K., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, *9*(1), 4787. https://doi.org/10.1038/s41467-018-06930-7

Starbird, K. (2019). Disinformation's spread: Bots, trolls and all of us. *Nature*, *571*(7766), 449.

Sundar, S. S., Molina, M. D., & Cho, E. (2021). Seeing is believing: Is video modality more powerful in spreading fake news via online messaging apps? *Journal of Computer-Mediated Communication*, *26*, 301–319. https://doi.org/10.1093/jcmc/zmab010

Taylor, D. M., & Doria, J. R. (1981). Self-serving and group-serving bias in attribution. *The Journal of Social Psychology*, *113*(2), 201–211. https://doi.org/10.1080/00224545.1981.9924371

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, *6*(1), 1–13. https://doi.org/10.1177/2056305120903408

Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C. H., Matthes, J., & Stanyer, J. (2017). Political communication in a high-choice media environment: A challenge for democracy? *Annals of the International Communication Association*, *4*, 3–27.

Van der Linden, S., Panagopoulos, C., Azevedo, F., & Jost, J. T. (2021). The paranoid style in American politics revisited: An ideological asymmetry in conspiratorial thinking. *Political Psychology*, *42*(1), 23–51. https://doi.org/10.1111/pops.12681

Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking. *Council of Europe Report*. http://tverezo.info/wp-content/uploads/2017/11/PREMS-162317-GBR-2018-Report-desinformation-A4-BAT.pdf

Waisbord, S. (2018). Truth is what happens to news. *Journalism Studies*, *19*(13), 1866–1878. https://doi.org/10.1080/1461670X.2018.1492881

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review*, *9*(11), 39–52. http://doi.org/10.22215/timreview/1282

Yang, C., Ding, L., Chen, Y., & Li, H. (2021, July). Defending against gan-based deepfake attacks via transformation-aware adversarial faces. In *2021 international joint conference on neural networks (IJCNN)* (pp. 1–8). IEEE. https://arxiv.org/abs/2006.07421

Zhang, J., Carpenter, D., & Ko, M. (2013). Online astroturfing: A theoretical perspective. *AMCIS 2013 proceedings*. https://www.researchgate.net/profile/Darrell-Carpenter/publication/286729041_Online_astroturfing_A_theoretical_perspective/links/56df195908ae979addef5103/Online-astroturfing-A-theoretical-perspective.pdf

Author biographies

Michael Hameleers (PhD, University of Amsterdam) is an assistant professor of Political Communication at the Amsterdam School of Communication Research (ASCoR). His research interests include (right-wing) populism, disinformation, and selective exposure.

Toni G. L. A. van der Meer is an assistant professor at the Department of Corporate Communication of ASCoR, University of Amsterdam. His research interests include crisis communication, (negativity) biases in the news, media literacy and misinformation.

Tom Dobber is a postdoctoral researcher at the Amsterdam School of Communication Research (ASCoR). His research focuses on political microtargeting and electoral pledges.